



Data harmonization of environmental variables: from simple to general solutions

O. Baume (1)

(1) Department of Environmental Sciences, Wageningen University, the Netherlands (olivier.baume @wur.nl), (2) European Commission - JRC, Land Management & Natural Hazards Unit, Ispra (VA), Italy (florence.carre@jrc.it)

European data platforms often contain measurements from different regional or national networks. As standards and protocols – e.g. type of measurement devices, sensors or measurement site classification, laboratory analysis and post-processing methods, vary between networks, discontinuities will appear when mapping the target variable at an international scale. Standardisation is generally a costly solution and does not allow classical statistical analysis of previously reported values. As an alternative, harmonization should be envisaged as an integrated step in mapping procedures across borders.

In this paper, several harmonization solutions developed under the INTAMAP FP6 project are presented. The INTAMAP FP6 project is currently developing an interoperable framework for real-time automatic mapping of critical environmental variables by extending spatial statistical methods to web-based implementations.

Harmonization is often considered as a pre-processing step in statistical data analysis workflow. If biases are assessed with little knowledge about the target variable – in particular when no explanatory covariate is integrated, a harmonization procedure along borders or between regionally overlapping networks may be adopted (Skøien et al., 2007).

In this case, bias is estimated as the systematic difference between line or local predictions.

On the other hand, when covariates can be included in spatial prediction, the harmonization step is integrated in the whole model estimation procedure, and, therefore, is no longer an independent pre-processing step of the automatic mapping process (Baume et al., 2007). In this case, bias factors become integrated parameters of the geostatistical model and are estimated alongside the other model parameters.

The harmonization methods developed within the INTAMAP project were first applied within the field of radiation, where the European Radiological Data Exchange Platform (EURDEP) – <http://eurdep.jrc.ec.europa.eu/> – has been active for all member states for more than a decade (de Cort and de Vries, 1997). This database contains biases because of the different networks processes used in data reporting (Bossew et al., 2007). In a comparison study, monthly averaged Gamma dose measurements from eight European countries were using the methods described above. Baume et al. (2008) showed that both methods yield similar results and can detect and remove bias from the EURDEP database.

To broaden the potential of the methods developed within the INTAMAP project, another application example taken from soil science is presented in this paper. The Carbon/Nitrogen (C/N) ratio of forest soils is one of the best predictors for evaluating soil functions such as used in climate change issues. Although soil samples were analyzed according to a common European laboratory method, Carré et al. (2008) concluded that systematic errors are introduced in the measurements due to calibration issues and instability of the sample. The application of the harmonization procedures showed that bias could be adequately removed, although the procedures have difficulty to distinguish real differences from bias.